No. 409
1956

Vol. 70
No. 2

A Study of Adoptive Children: II. The Predictive Validity of the Yale Developmental Examination of Infant Behavior

By

J. R. Wittenborn and Others

Rutgers University

Vol. 70, No. 2 Whole No. 409, 1956

Psychological Monographs: General and Applied

A Study of Adoptive Children: II. The Predictive Validity of the Yale Developmental Examination of Infant Behavior[1]

J. R. Wittenborn
Rutgers University

With the collaboration of:

Myrtle A. Astrachan, Ph.D., Psychologist
Marjorie W. DeGooyer, M.S., Psychologist
W. Wallace Grant, M.D., Pediatrician
Irma Z. Janoff, Ph.D., Psychologist
Robert B. Kugel, M.D., Pediatrician
Barbara J. Myers, M.S., Psychologist
Annaliese Riess, Ph.D., Psychologist
Ellery C. Russell, M.S., Social Worker

Introduction

The examination developed at the Yale Clinic under the pioneering leadership of Dr. Arnold Gesell has become widely known and in many areas is used as the principal aid in the placement of adoptive children. Despite its wide acceptance and important role in many practical situations, the validity of this infant examination has never been explored by suitable investigation. Since Cattell (5),[2] Anderson (1), and Bayley (3) have been unable to find any evidence of useful validity for their infant examinations, some reasonable doubt concerning the validity of the Yale examination has arisen.

There are at least two possible interpretations of a general question concerning the validity of an infant examination. One interpretation concerns the question of whether or not the infant examination provides a relevant and acceptable evaluation of developmental status at the time of the infant examination. Gesell has commented on this aspect of validity in the following manner:

> Fundamentally the validity of the schedule here offered depends on the validity of the norms, the legitimacy of the category classi-

[1] The staff of the project wishes to express its appreciation to Dr. Donald Young, Director of the Russell Sage Foundation, and to Dr. Carl Hovland and Dr. Milton Senn of Yale University. The nature and the quality of this effort have been greatly influenced by such friends and advisers as Dr. Joseph Baldwin, Dr. Nancy Bayley, Dr. Harold Jones, Dr. Neal Miller, Dr. Robert Sears, and Dr. Helen Thompson. We are indebted also to Dr. Leonard Cottrell, Dr. Liselotte Fischer, Dr. Else Frenkel-Brunswik, Dr. Marjorie Honzik, Dr. Mary Cover Jones, Dr. Frederick Mosteller, and Dr. A. J. J. Van Woerkom. We are particularly grateful to the adoptive families who participated in this investigation, the child placement agencies who cooperated with us and guided our efforts, and to Adelaide Piscitelli, David Macklin, David Jenkins, and Peter Field who helped in the analysis of the data and the preparation of the manuscript.

[2] References for the present report are listed on page 114. The pagination of the present report is continuous with the first report of this series, entitled "A study of adoptive children: I. Interviews as a source of scores for children and their homes," *Psychol. Monogr.*, 1956, Vol. 70, No. 1 (Whole No. 409).

fications, the appropriateness of each item for the category to which it is allocated, the soundness of the concept of maturity level, and the justness of using a sample of the child's behavior to indicate that level . . . (7, p. 219).

Our investigation is not concerned with validity of the infant examination as a basis for evaluating the developmental *status* of an infant.

Our investigation is concerned with a second aspect of validity. What are the predictive implications of the knowledge of the current developmental status of an infant provided by the infant examination? Gesell and his associates may also be quoted with respect to the aspect of validity:

> How consistent over a decade and more have been these mental-growth careers? In no instance did the course of growth prove whimsical or erratic. The behavior biographies give clear evidence of a high degree of latent predictability, even in infancy (6, p. 49).

> One of the greatest advances that we know of in adoption practice in the last twenty years has been the policy of giving careful developmental or psychological examinations to every single adoption candidate. Such examinations indicate whether an infant or child is of "normal" mentality, and therefore a good adoption candidate, or if he is so subnormal that he is not a fit candidate. They further can to some extent discriminate within the normal group children among average, above-average, and below-average endowment—thus giving a good clue as to what kind of home each child would best fit (10).

It is this predictive aspect of validity that provides the major question of the present investigation.

Our study of the predictive validity of the infant examination required measures for three classes of information. One class of information emerged from the infant examination; the second class of information described the child at the time of the follow-up study; the third class of information described some relevant facets of the child's environment.

It should be emphasized at the very beginning that a validity study is meaningful only to the extent that it shows evidence of relationship between the device to be validated and some criterion. Specifically, the most informative result of our investigation could emerge in those instances where our analyses reveal correlations between some aspect of the infant examination and some aspect of our criteria of development. Absence of relationship can be due to unreliability in the criterion, restrictions or inconsistencies in the sample, and numerous other factors, including limitations of the measuring instrument which is to be validated. Such limitations include lack of intrinsic validity for the criterion under consideration and lack of sufficient reliability to enable the instrument to be meaningfully related with any criterion.

Because of the ambiguous status of lack of relationships between the infant examination and criteria for development, we have sought to use all possible facets of the infant examination, all available measurements of childhood development, and all defensible subsamples of our adoptive children in an effort to show evidence of predictive validity. We may state this differently by saying that we have sought exhaustively to disprove the hypothesis of no predictive validity for the infant examination. As a matter of fact, this is the only defensible rationale in such an investigation because one can never prove an hypothesis of no relationship.

In the following paragraphs it will be noted that we have used several samples and subsamples in our exploration of the predictive possibilities of the infant examination. Since some of our adoptive children were placed in their adoptive homes selectively as a result of the infant examination (evidence for this will

be provided later in this report), we have provided separate analyses for children who were placed in their adoptive homes *before* the infant examination and where presumably environmental differences resulting from selective placement could not be confounded with relationships between the infant examination and criteria for development.

Children who have been placed adoptively *after* the infant examination may not be ideal subjects for a validative study such as ours. Trends emerging for such children are suspect because they could occur as a consequence of confounding with environmental differences occurring as a result of selective placement. Such selective placement could result either from knowledge of true parents of the infants or from knowledge of the infants' standing on the infant examination. With respect to the danger of confounding as a result of knowledge of the true parents, we found a group of nonadoptive children living in their own homes; they provide a very desirable control. With respect to the possibility of confounding because of selective placement based on knowledge of the infant examination, we found certain measures of the environment to provide us with a desired control and means for analyses relevant to this particular selective placement hypothesis.

Because of numerous complexities introduced as a result of the practical nature of our investigation and as a result of the placement practices of social workers in southern Connecticut during the middle forties, some of our analyses are relatively easy to describe and others are relatively difficult to describe. Accordingly, we shall use simple subgroups in presenting the essential nature of our analyses. In our conclusions, we shall summarize the implications of our analyses for all the various measures and the various subgroups combined.

Since some of our discussions will be based upon a comparison of the relationships emerging from different subgroups, one might suspect that such differences are not due to the variables which we have in mind but are an artifact of some differences between subgroups such as range in test scores or such as age at the time the infant examination was administered, etc. It shall be apparent in our discussion of our subsamples that we have given proper consideration to these possibilities.

The Infant Examination

The infant examination developed by Gesell and his associates is not a psychometric test. It is a procedure for clinical examination and yields several kinds of descriptions and evaluations (7). Accordingly, any effort to evaluate the procedure must be concerned with its various facets and must attempt to validate them independently. In order best to describe the various facets, it would seem desirable first to describe the examination procedure in general. Our description will be brief, however, because authoritative publications are available which describe the examination fully, including the conditions under which the examination is given, the exact manner in which it is scored, and the precise nature of the norms employed (7, 8).

The infant examination procedure is based on observation of the infant's responses to a series of standard situations. These include response to the examiner, response to objects and situations, and in addition, certain reported responses contributed by the child's mother or nurse. The range of responses included in the examination procedure varies with the age of the child and with his developmental status:

it is unlikely that infants who differ from each other by several months in age are evaluated with respect to precisely the same kinds of responses. The infant examination procedure, like the Binet scales, changes in content for children of different ages. The opinion of many who employ the examination suggests that it may have its greatest value when administered to children whose ages lie between 12 and 40 weeks. (Some consider this to be due to the developmental pattern of infants; others consider it to be due to limitations in the content of the examination. Incidentally many practitioners claim that the examination may be used more reliably for identifying children who have a questionable developmental status than for distinguishing between children of superior developmental status.)

The examination yields part scores. These are available on the basis of an a priori classification of the items. This classification includes: fine motor, gross motor, linguistic, personal-social, and adaptive aspects of infant development. It is also common practice to combine the child's performance on these five parts of the examination procedure in order to provide a general estimate of infant development. It is customary to express the infant's developmental status with respect to developmental age equivalents. Such age equivalents are based on norms constructed from previous observations of children. It is possible, of course, to convert these developmental age equivalents to developmental quotients; this arithmetic adjustment for age, however, appears not to be a general clinical practice.

Although the examination provides a standard situation for observing the child, one finds diverse supplementary descriptions in the protocols. For example, in the examiner's written description of the child such words as mature, friendly, weak, etc., are often used. Some of these words occur with sufficient frequency in a set of protocols to justify their informal inclusion as an aspect or a part of the results of the examination procedure. In addition to such informal characterizations of the child and his responses, there are other descriptions that are more formal in their nature and have definite implications. For example, there are often diagnostic statements included with the examiner's description of the child, and some of these refer to the child's physiological or neurological status.

A principal feature of the examination is the general classification for the infant, a classification based upon a combination of the aspects of the child's development which are expressible formally in terms of age equivalent scores, the subjective impressions of the examiner, and various qualifying features which make the distinction between a clinical evaluation and a psychometric test. The classification of the infant was established with respect to levels, and included superior, high average, average, low average, dull, and sometimes such classifications as borderline and moron. It should be noted at the outset that in the case of many examinations no classification was given for the child. This ordinarily meant that for some reason or other the examiner was unwilling to commit himself with respect to the child's status, i.e., for some reason or other he considered the examination unsatisfactory or undependable. Sometimes this was due to some temporary indisposition on the part of the child. Occasionally, moreover, it occurred when the child's performance was definitely unsatisfactory and, if accepted at face value, would result in a very low or unfavorable classification. It is to the credit of the clinic and its individual examiners that they were extraordinarily cautious in classifying children whose performance was inferior. The procedure in questionable cases was to request further examination of the child.

The results of the examination were also expressed with respect to recommendations concerning the adoptability of the child (assuming that he was not already adopted). If, in the examination procedure, the child performed well (this usually was a basis for a confident examination), he was recommended as adoptable. Sometimes the recommendation was qualified and in some cases he was considered not adoptable. The recommendation of "not adoptable" was usually further qualified with the recommendation that the child be evaluated again at some later time. Other recommendations were made concerning the suitability of his current boarding or other temporary placement. For some cases it was suggested that the child was overplaced in the sense that the boarding home was too demanding or too stimulating for him; in other cases it was suggested that he was underplaced and that he could grow more rapidly in a stimulating home. It is difficult to determine in most cases how well the specific rec-

ommendations of the clinic were followed by the agencies or individuals requesting the examination.

It is probable that an examination would, in the case of many children, determine the immediate course of action (not necessarily the action recommended by the clinic, however). This seems particularly likely when one considers the conditions under which the infant test was given. An infant test could be initiated at the request of an adoptive parent, at the request of an agency handling the adoption, or at the recommendation of the court. Inasmuch as some one individual or a group of individuals had to take the initiative in securing an examination, it seems most likely that the results of such an examination would be used as an aid in planning the next step in the handling of the child. This possibility is anticipated in the introduction to the present report and provides one of the most obvious and also one of the most difficult complications in the present study.

It may be re-emphasized that the infant examination comprises many facets and that we should seek evidence for the validity of each of these various facets. They include the characterizations of the child, special diagnoses, developmental status with regard to the various parts of the examination and to the total examination; a general classification of the child based on performance on items and on general impression; recommendations with regard to adoptability, foster home, and other current circumstances. In some important respects the examination procedure was used cautiously; specifically, if a child did not perform well on the infant examination, subsequent examination was requested for him, classification may have been deferred, and presumably the possibility of immediate favorable placement was reduced. This means that concern was shown by the clinic for the possible untoward consequence of the spuriously low classification. It should be noted, however, that although the clinic itself tended to withhold classification until a satisfactory examination had been completed, it could not very well eliminate the possibility that a decision concerning the child would be made as a result of the first examination whether or not this examination were a satisfactory one from the standpoint of the clinic. If, for example, the outlook were unfavorable, this would be communicated to the agency and might prejudice their reaction to the child. At the very least, the qualified impression of the examiner would reduce the possibilities for a confident favorable placement of the child.

Since a second examination usually occurred as a consequence of a doubtful first examination, it is probable that children with more than one examination were classified more confidently on the final examination than they were on the first. It is not necessary, however, that a final examination should have a greater intrinsic validity than the first. On the contrary, there is some possibility that the results of the first examination could be prejudicial to the result of a later examination. Even if this possibility could rarely or never be demonstrably attributed to subjectivity among the examiners, it nevertheless seems quite possible that a qualified or unfavorable evaluation of the child at the time of the first examination could in many cases reduce the possibility of his having optimal opportunities for development. A qualified performance on the

first examination could, in effect, perpetuate the kinds of conditions that existed at the time of the first examination so that the possibility of change in development on the part of the child could be restricted. Thus, it not only seems possible that the relationship between the final infant examination and some later criterion for development could be confounded by opinions based on the first examination, but that the relationship between the first examination and a final infant examination could also be confounded by a perpetuation of the same kind of factors that influenced performance on the first.

Description of Subjects

Since our sample of subjects was drawn from children who had been examined in infancy at the Yale Clinic of Child Development, many important properties of our sample are determined by the kinds of cases to which the infant examination had been applied at the clinic. Our survey of the records there indicated that during the period beginning in 1942 and ending in 1947 the children to whom the infant examination had been administered fell into three major groups (only a few were exceptional to this grouping).

One group comprised children who were ill, defective, or in some respect a cause of concern to their parents. Almost all of these children were living with their true parents, and none of them was being considered for adoptive placement. They were suffering from all sorts of disorders, many of them diagnosable as Mongolism, hydrocephaly, microcephaly, cerebral palsy, convulsions, birth injury, etc. Although the infant examination may have contributed to a refined or confirmed diagnosis for these children or to an exact evaluation of their current developmental status, this group of children was brought to the clinic for examination because of apparent or suspected pathology. As far as we know, none of the infants in this group was brought in for examination because at that time he was being considered for adoptive placement or because there was some question as to whether or not he should soon be considered for adoptive placement.

The second group of infants brought in during this period comprised those who were being considered for adoptive placement. The infant examinations were desired by child placement agencies, or by courts, or by adoptive parents. Although many of these children had a developmental status which was very modest if not questionable[3] and although some of them were weak and possibly ill, all of these children were being considered as adoptive candidates. It would appear to us that the children being considered for adoptive placement and the children being considered for clinical evaluation were quite generally and consistently different from each other so that the discernible defects that characterized the clinical group were relatively inconspicuous among the group being considered for adoption. Apparently very few grossly ill or defective children were being considered for adoptive placement during this period.

There was a third group of children examined during this period. They were

[3] It may be noted that for our younger sample the mean and standard deviation of the General Maturity Quotient on the infant examination were 96.60 and 15.77, respectively. Only one per cent of these children had General Maturity Quotients below 60, 15 per cent were 60 or above but below 80, 33 per cent were 80 or above but below 100, 46 per cent were 100 or above but below 120, and 5 per cent were 120 or above.

living with their true parents and were examined as a part of the study of prospective members of the Yale Clinic Nursery School. This group was a relatively favored group; many of their parents were graduate students or members of the Yale faculty.

In our investigation, we drew our principal samples from the children who had been considered for adoptive placement. This means that very few, if any, grossly defective infants contribute to our results concerning the predictive validity of the Yale Developmental Examination. One should not generalize the results of our investigation to a possible use of the infant examination as a means of distinguishing between grossly defective children who are not adoptable by usual standards and questionably adoptable children whose development is slow or may have been retarded. Since the grossly defective or ill children did not appear in the group being considered for adoptive placement, one could suppose that for the most part these children are identifiable without the aid of the infant examination. Accordingly, as the infant examination was used in the Yale Clinic of Child Development during this period, its value as an aid in adoptive placement appeared not to rest on its value in making distinctions between the obviously ill or grossly defective and the infants of questionable development. We shall attempt to validate the use of the infant examination as an aid in the evaluation of infants being considered for adoptive placement.

Since the proper use of the infant examination may depend greatly upon the skill of the examiner, and since it is possible that, to an important degree at least, infant examinations are a standard medium for the skillful clinician to make his observations, it was decided that a study of the predictive validity of the examination should be based on the protocols of the *most approved* examiners.

Our sample of subjects is also influenced by the fact that we were interested in the predictive validity of the examination when applied to infants, particularly adoptive infants. As a consequence, we excluded from consideration all children who did not receive their first examination before they were 14 months of age.[4] Thus we are studying the predictive validity of an *infant* examination; we are not studying the predictive validity of an examination for very young children.

The records indicate that the examiners of the Yale Clinic of Child Development were very cautious in their evaluations and recommendations. It is amply evident that they sought to protect the adoptive children from a negative evaluation which could have been spuriously low or could have had spuriously negative implications. It would appear that they sought also to protect the adoptive parents from adoptive infants of questionable developmental potential. Because of these precautions, many of the infants were not promptly classified by the examiners. Instead, it was recommended that the infant be returned at some later time for further examination. It was further recommended that adoptive procedures be delayed until a more confident evaluation of the infant was possible. Accordingly, our samples could be seen as comprising children who had been confidently classified at the time of their first infant examination and children who had not, and were in turn

[4] Much of our discussion of the infant examination will refer explicitly to subgroups who were between 12 and 40 weeks of age at the time of the infant examination.

subject to further examination. We made independent studies of the predictive implications of the infant examination for subgroups of children who differed with respect to the number of infant examinations they received. The various subgroups provided the same implications with respect to the question of the predictive validity of the infant examination. As could be imagined, however, the analyses for the children who had but one infant examination were the most simple; accordingly these are the analyses that receive a particular emphasis in the later sections. These analyses have the additional merit of being based on children whose first infant examination was considered to be most dependable.

One major characteristic of the sample was determined arbitrarily by us and had little to do with the testing practices of the clinic. This was the question of the age of the children at the time of our follow-up inquiry. In general, we wished to base our study on children who were as old as possible. In this way we could have not only knowledge of the children over a relatively long period of time, but we could also take advantage of the fact that there are more satisfactory measuring devices available for older children than there are for relatively young children. There were too few children in adolescence at the time of our investigation to enable us to make a satisfactory study of this stage of life. We believed that children in the first grade were in many cases going through a kind of turmoil or a period of uncertainty and rapid change, so that they would not be ideal children for a predictive study. This left us with two groups of children from which to choose; the five-year-olds who were old enough to test in a reliable manner, but not yet in the throes of the first grade, and the seven-, eight-, and nine-year-olds who were past the throes of the first grade but not yet involved in the preadolescent changes which assume such remarkably different patterns in different children. Sound arguments could be advanced favoring either the preschool group or the school group as the more appropriate for examining the predictive validity of infant examinations. Because of this, we decided to employ both samples, and we hope that the use of two samples of different ages will enable us to provide an element of replication in our study so that the implications for one sample can serve to confirm or to challenge the implications for the other sample. One hundred fourteen children were included in our younger sample, and 81 in our older sample.

We began our investigation by arranging a roster of adoptive children who had been examined before 14 months of age by approved examiners and who at the time of our follow-up inquiry would be at the proper age for either our younger or our older sample. In order to explore the nature and the degree of selective factors in our attempts to reach adoptive children for further study, we selected 170 names from our roster of younger children and sought intensively to include all of these 170 children in our study. The adoptive arrangements for 40% of these were handled independently and the adoptive arrangements for 60% of them were handled by a child service agency. It is interesting that we succeeded in including 56% of each of these groups in our study. (The balance of our younger sample of 114 and our older sample of 81 was secured in a similar way, except that accessibility became an important factor and relatively less effort was made to reach some

of the cases who were distant, uncertain, or in some other respect presented complications.) It should be remarked that the families whose child had been placed independently were found to be cooperative with our study wherever we were able to reach them. The independent families who did not participate in our study were without known address, living in some distant part, or willing to disregard our inquiry.

Since we undertook our investigation in the expectation of providing research information that would be of value to child placement agencies, we sought to work closely with these child placement agencies, and tended to regard ourselves as their research staff. In the case of all agency placements, we called on the responsible agency, described our research plan to them, and asked them to help us to secure the interest and cooperation of the adoptive families. It is obvious that our agency placements were screened in some way or other by the agency. We cannot be sure how this may have affected the outcome of our study but it certainly did not prevent us from studying many families where the adoptive situation was uncomfortable either because of probable limitations in the home or in the child.

In any study such as the present one, one must be aware of the possibility that the children available for follow-up inquiry are in some important respect different from those not available, and as a consequence, the results of the investigation are in some way qualified and their general applicability limited. Because of this we made elaborate comparisons (with respect to infant records) between the younger children included and those not included in our follow-up study.

Our total roster of younger adoptive children who had been examined in infancy before 14 months of age and whose examinations were conducted by approved persons comprised 310 names. As indicated in an earlier paragraph, we made an intensive attempt to include a group of 170 of these children in our sample, and succeeded in including 95 of them. An additional 19 younger children were included on a "first come, first served" basis. Thus our study of the younger children is based on a total group of 114. Accordingly, our comparison between those available for a follow-up study and those not available for follow-up study involves the 114 younger children we had seen and the 196 younger children we had not seen.

With respect to the kinds of information that were descriptive of the infant, the results of our comparisons are reassuring. There was no reliable difference between the children we saw and the ones we did not see with regard to such factors as prematurity at birth, complications at birth, physical deviations, age at leaving the hospital, general maturity quotient on the infant examination, classification on the infant examination, outlook on the infant examination, examiner's recommendations concerning adoptability, and advice concerning the care of the child at time of infant examination. Furthermore, there were no differences with regard to such informal designations as active, attractive, attentive, sustained drive and interest, and stable. There were two such informal designations, *socially responsive* and *mature,* which offered some evidence of differences between the children in the follow-up group and those not in the follow-up group.[5]

[5] Since so few of the numerous comparisons show a trend, we may suspect that these trends arose by chance.

The reader may suspect that the children included in our study were different from those not included in our study with respect to the attainments or circumstances of the true parents. Although our data were not complete, we made comparisons between those seen and those not seen, with respect to such factors as the education and the occupation of the true father, and the education, occupation, and health history of the true mother. With regard to these classes of information, there was only one indication of a reliable difference; it was found that the true fathers of the children not included in our study had slightly, but reliably, higher educational attainments than the fathers of the children included in our follow-up study.

With respect to the kinds of early information which are not descriptive of the infant per se, but are descriptive of his care and general circumstances, we find that our comparisons yield some reliable trends. The children who went into an adoptive home after their first preliminary placement were somewhat more available to us than the others. This may suggest that adoptive parents who receive their children relatively early are more confident and cooperative about adoptive matters, including follow-up studies, than are parents who got their children after several preliminary placements. The examiner appears to have had some effect on the availability of the children for our study. We had learned in our interviews with the adoptive parents that in some cases the parents were in strong disagreement with the examiner and were displeased. Possibly displeasure arising from the examiner could have made some of our adoptive parents unwilling to cooperate in the follow-up study. This possible difference in the examiner cannot be shown to be due to intrinsic differences in the child (so far as we could discover, there was no practice in the Clinic that would involve giving all the "tough" cases to one examiner, etc.). Concerning the implications of differences among examiners, we have found that those examiners whose cases were underrepresented in our follow-up study tended in general to be examiners whose infants had general maturity quotients that were relatively low. Some examiners either drew relatively inferior children or were relatively pessimistic in evaluations of their children. Such differences among examiners are consistent from one test to another.

With respect to the data which are available to us, at least, there appear to be no important differences between children that we were able to include and those we were unable to include in our study. At least, any differences which our comparisons suggest cannot be seen to qualify the implications of our analysis.

Let us summarize our sampling procedure and provide a brief description of our samples. Our study is concerned with two major samples of adoptive children. There was one group of 114 *younger* children, most of whom were in their fifth or sixth year and none of whom was involved in the first grade. There was a second group of 81 *older* children. Most of these children were in their eighth or ninth year and none of them was in the first grade. Agency and independent adoptive placements were represented in both our younger and older sample, but not in equal proportions. Forty per cent of the younger sample (i.e., those more recently placed) were handled independently of any child placement agency, whereas 20% of the older sample were independent place-

ments. Almost all of the children who had been placed by an agency were placed in their adoptive homes *after* they had received an infant examination, and we may suspect that the results of the infant examination may have been a factor in the agency's selection of the adoptive home. Almost none of the independent placements were placed in adoptive homes *after* they received an infant examination; accordingly, we have no reason for suspecting that the infant examination could have been generally important in the adoptive placement of the children who were handled independently.

Our report makes incidental use of a third sample of 35 children who live with their true parents and who were examined in anticipation of their attendance at the Clinic nursery school. This group was employed primarily as a control group in our study.

In the course of our validity analyses, it was necessary for us to make two kinds of distinctions among our children. First, we were interested in whether or not the child was placed in the adoptive home before the infant examination; this is relevant to the possibility of confounding. Second, we were interested in whether or not the first infant examination was optimal from the standpoint of possible validity; specifically, we were interested in whether or not the first examination permitted the examiner to classify the infant promptly and whether or not it was made at the best possible age to show validity. (Experts had advised us that the Yale infant examination may be most valid when it is given between the ages of twelve and forty weeks. It was indicated that infants under twelve weeks were too young to show a repertory of behavior that would permit valid evaluation, and it was indicated also that the growth norms for the items comprising the examination procedure were of such a nature as to make distinctions between older infants relatively inexact, certainly less exact than those possible for infants under forty weeks. This presumptive feature of the growth norms is congruent with the tables published by Gesell and Thompson [8].)

Table 1 provides a summary of our sample of adoptive children and a classification from the standpoint of relevant subsamples. It may be noted that 62 of our younger children received their first infant examinations at what was considered an optimal age and were classified as a result of the examination. Our division of these children into the 32 who were placed in their adoptive homes *before* the infant examination and the 30 who were placed in their adoptive homes *after* the infant examination is explicitly involved in one phase of our analysis and we have referred to these as the selected "before" and the selected "after" groups, respectively. Since we shall be involved in certain comparisons between these two groups, it should be noted that the means and the variabilities of these two groups are comparable with each other, not only with respect to age at the time of the first infant examination, but with respect to status on the first infant examination, later Binet IQ, and other relevant data.

Criteria of Development

There are many published reports based on the Yale Developmental Examination of Infant Behavior. Among them various statements may be found which imply that the examination procedure has a predictive significance for numerous different aspects of child development. So far as the writer is aware, however, there is no published statement

TABLE 1

A CLASSIFICATION OF THE CHILDREN WHO PARTICIPATED IN THE STUDY

Sample	The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks*	The first infant examination did not result in a confident classification and/or examinations were administered after 40 weeks of age**	Total
I. THE YOUNGER ADOPTIVE SAMPLE:			
A. Infants placed in adoptive homes *before* the infant examination	32	22***	54
B. Infants placed in adoptive homes *after* the infant examination	30	30	60
Total	**62**	**52**	**114**
II. THE OLDER ADOPTIVE SAMPLE:			
A. Infants placed in adoptive homes *before* the infant examination	16	10	26
B. Infants placed in adoptive homes *after* the infant examination	37	18	55
Total	**53**	**28**	**81**
III. THE NURSERY SCHOOL SAMPLE (normal children living with their true parents and examined in infancy as part of plans for the Clinic Nursery School)			35

* Experts had advised us that the infant examination may be most valid when given in the 12 to 40 week range.
** Many of the children were not placed in adoptive homes until they had had two or more examinations.
*** Two of these were examined before 12 weeks; one at 9 weeks, the other at 11 weeks of age.

specifying the criteria appropriate for validating any of the possible predictive implications of the infant examination. This creates a difficult situation for one who seeks evidence for the predictive validity of the infant examination. If one were to be narrowly selective in deciding upon appropriate criteria, one would necessarily run the risk of omitting aspects of development which in the minds of some interested persons could be the basis for the most relevant of criteria. If, however, one wishes broadly to explore the predictive implications of the infant examination, two difficulties must be anticipated. One of these is the difficulty which is necessarily associated with gathering data relevant to numerous and diverse criteria of child development. The second difficulty emerges from the fact that standardized, generally applicable measures have not been developed for the personal and social development of children. For the most part, the devices available for studying the personal and social development of children are clinical procedures that are not only unvalidated but are of such a nature as to make validation most difficult.

Since we sought evidence for the predictive validity of the infant examination, we decided to measure and assess our children in all *feasible* respects. As we have indicated elsewhere, however, the fact that our children are adoptive children and that their availability to us is a result of the voluntary cooperation of their parents greatly qualifies what is *feasible*. Our examination procedure must be not only extraordinarily inoffensive and efficacious, but it must respect the tender status of adoptive children in our society.

In the following paragraphs we shall

describe the various criteria that were available to us. Some of the criteria, such as measures of intelligence, are tried and true; others, such as some of the scores emerging from the mother's description of the child are novel, untried, and without evidence of a high degree of reliability. With regard to the novel and untried criteria which we shall employ, we doubt if they should be depended upon to indicate the degree or precision with which such criteria might be predicted by the infant examination. Nevertheless, we believe that they possess sufficient reliability to satisfy a useful purpose. Specifically, if the infant examination has a predictive validity for the aspects of behavior represented by our novel criteria, we believe that the criteria we use will be sufficient to show this possibility.

In general, we hope to show evidence that the infant examination has predictive validity for the aspects of development which may be represented in our investigation. Since we use numerous criteria, it is not implied that we consider the infant examination committed to the prediction of all such criteria as these. In effect, we shall proceed to test a series of null hypotheses. Specifically, we shall proceed in the manner which is conventional to all modern investigations concerning the testing of hypothesized relationships between variables. We shall think in terms of a series of null hypotheses, all of which could be made explicit in such a form as "there is no relationship between [the specified aspect of the infant examination] and [the specified criterion for development]." We shall then proceed with our analyses in a manner calculated to refute as many as possible of the null hypotheses. It is probably unnecessary to remind the reader that a null hypothesis is a hypothesis to be *disproved,* and that it is illogical to presume to *prove* a null hypothesis or to establish its truth. Accordingly, our most meaningful results occur in those instances where we can refute a null hypothesis, because wherever we can refute it, the hypothesis that the infant examination has predictive validity is unchallenged and by implication could be true. Accordingly, wherever the null hypothesis is refuted, we may become interested in the degree of relationship between the aspect of the infant examination and the criterion of development that is revealed in our sample, and for our tried and true criteria at least, we may be interested in contemplating the predictive economy that could result from the use of an infant examination.

Failures to refute null hypotheses can be due to many causes, such as the use of criteria that have an inappropriate content, the use of criteria that have too little reliability to show a significant trend, the use of samples that are restricted or are in some other way inappropriate; and of course, it is possible that the predictor variable simply has no predictive value for any of the criteria under consideration.

There are doubtless many students of child development who would say, "The status of many features of child development is not clear; how can you arbitrarily specify what is good development and what is poor development in a child?" Because of the controversial status of some features of child development, we have eliminated them from consideration. Most of the aspects of child development which we have included in our criteria are ones with respect to which a large number of people would be in agreement, i.e., they would feel that they understood the relevance

of the criteria for some of the values of our society. Our criteria may be outlined as follows:

I. *Intellectual Development*

A. The Stanford-Binet, 1937 Revision (12). The Binet test is considered by most people to be a good measure of the child's mental development, particularly with respect to those features of the mental development that represent the acquisition of verbal skills or that depend upon employment of verbal skills for their expression.

B. The Arthur Point Performance Scale (2). A performance scale was included because it is possible that the acquisition of verbal skills is stimulated more by certain homes, whereas the acquisition of nonverbal abilities, e.g., spatial understanding and manipulating skills, is stimulated more by other kinds of homes. The Arthur was selected because it provides a comprehensive sample of important nonverbal abilities and because it has norms suitable for a relatively wide age range.

II. *Scholastic*

The possibility had been suggested that the developmental propensities measured best by the infant test may be brought out and made distinct by school experiences. It was decided, therefore, to employ two samples. One sample would comprise children who were old enough to be cooperative and responsive in a formal testing situation but not yet old enough to be in school; another sample would comprise children who were in school and had passed the throes of the first grade. Since it was decided that some of our children should be in school at the time of follow-up, a useful measure of their development would be some estimate of their scholastic achievements. We have used measures for reading, spelling, and arithmetic. The measures we employed (Gray Oral Reading Paragraphs Test, Stanford Achievement Tests Spelling List—Form D, and Stanford Achievement Test Primary Arithmetic Test—Form D) are standardized and widely accepted as appropriate evidences of achievement at the grammar school level (9). Data for scholastic achievement are, of course, not available for our younger subjects.

III. *Motor Development*

Motor development of our children was sampled by employment of two tests of strength—jumping and a hand dynamometer test. In addition, the proficiency with which children use their hands in the Arthur examination was rated. Ratings of alacrity and coordination were also made by the examining physician as he observed the child during the course of the pediatric examination. The strength tests were administered as part of the pediatric examination.

IV. *General Physical Development*

The general physical development of the children was appraised by collaborating pediatricians who were Fellows in the Yale Child Study Center. The doctors weighed the children; took their height; observed them during the course of a general pediatric examination which included an examination of the child's development, his sensory acuity, and his reflexive responses. The physical examination, although in general quite thorough, did not include an intensive study of any particular facet of the child's development. The examination was designed to reveal discernible, practically significant deviations in physical development of functions. The examination was

not organized to reveal obscure or rare defects of structure or function.[6]

V. *Personal Development*

The aspects of the child's development most difficult to evaluate are those commonly designated as social or personal. Most of the devices that are available for evaluating a child's social or personal development have been prepared in clinical situations for the intensive evaluation of a given individual. They are not of such a nature as to permit the gathering of standard information for a large number of children. Because of this, the usual devices for studying children were not available to us, e.g., the Thematic Apperception Test or the Rorschach Test. It is only fair to point out that such well-known procedures are not only relatively idiosyncratic in their orientation but they themselves have not been validated and might indeed be more difficult to validate than the infant examination procedure per se. It seemed to us, therefore, that if it were necessary for us to use unvalidated criteria in order to examine certain aspects of the infant examination procedure, criteria should be adopted that have a certain amount of surface validity and that might possibly be subjected to independent validating studies. Evidence concerning the child's personal-social development is of three kinds:

A. An effort was made to sample the child's reaction to himself. In order to do this, we secured the children's responses to a series of hypothetical situations, for the most part situations familiar to them. We called this the Social Reaction Interview.

The information that we have gathered comprises many bits. It obviously would be undesirable to use each of these bits as one of our criteria. It seemed in the best interests of economy and clarity to combine or integrate these bits of information so that a conveniently limited number of variables of general meaning would be available.

The procedure for combining bits of information was very simple. All of the principal answers to each of the 111 questions in the Social Reaction Interview were correlated with the principal answer to every other question. Groups of questions which gave statistically consistent answers were formed. These groups were then refined by excluding any items that appeared to be psychologically inconsistent with the group. In this way groups or clusters of bits of information were formed that were both statistically and psychologically consistent. Any child could be given a score for a cluster simply by counting the number of answers he gave which fell in a particular cluster.

After such clusters were established, it was possible, for each of our children, to score the Social Reaction questions on the basis of each of these cluster scores. Such scores express the degree to which the kind of behavior implied by a cluster is characteristic of a child. Perhaps it is best to describe these scores as extensive scores rather than intensive scores. They do not imply how intense or severe any one response of a child might be, but instead indicate how general or pervasive a given kind of response might be for each child.

The manner in which this interview was used and scored is described in detail in *Interviews as a Source of Scores for Children and Their Homes* (See Part I of "A study of adoptive children"). Appendix A of that report includes a copy of the interview form which was

[6] The form on which the results of this examination were recorded is included in the Appendix to this monograph.

used. The different scores which emerged from this interview are listed below:

	Name of Cluster	Split-half Reliability
Younger sample		
	Dependence on Adults	.88
	Aggression	.58
	Socialized Compliance	.70
	Takes Adult Role	.69
	Weakness-Avoidance	.57
	Constructive Approach	.58
Older sample		
	"Goody-Goody"	.59
	Responsible Attitude	.52
	Cooperation with Authority	.41
Younger sample		
	Compulsive	.52
	Aggressive	.70
	Phobic	.57
	Dependent	.71
Older sample		
	Controlled	.78
	Spoiled Child	.78
	Dependent	.60
	Anxious-Aggressive	.80

It will be noted that the procedure yielded more scorable clusters for the younger sample than for the older sample. We consider this to be reasonable because the increased self-awareness of older children introduces complexities which greatly obscure the value of such a direct, explicit question-and-answer procedure.

B. While the children were being examined, the mothers participated in a semistructured interview with one of the psychologists. The major portion of this interview had to do with child-rearing practices. The content of this interview was recorded in standard form and organized with respect to specific bits of information. Many of these bits of information were descriptive of the child as seen by the mother and provided the basis for a cluster-scoring procedure. This interview and the cluster-scoring procedure are described in *Interviews as a Source of Scores for Children and Their Homes* (See "A study of adoptive children: I"), and the standard forms we used for recording the interview are included in Appendix B of that report. The following list includes the cluster scores that were descriptive of the children:

It should be noted that the nature of the cluster scores appears to shift with the age of the children. This we consider to be consistent with generally observable changes in children of different ages and this is described in the report mentioned earlier in this paragraph.

C. During the course of the psychological and pediatric examinations the child was carefully observed and ratings of his behavior were made during that time. These ratings obviously are based on a restricted sample of behavior. Nevertheless, how a child behaves in response to a challenging situation is considered by many people to be relevant to his social and personal development. (Samples of the rating forms are included in Appendix A of the present report.)

It is obvious to the reader that we do not have any measure of the whole child and indeed we have not sought such measures; research for the most part is concerned with abstractions or generalizations and, in this particular sense, is different from clinical responsibility which must ever be concerned with the individual child as a total pattern of responses.

It will be seen that our criteria are diversified rather than intensive. We have not sought to examine the child exhaustively with respect to any particular aspect of behavior; we have sought instead to sample many of the different aspects of behavior which conceivably

are relevant to the implications of an infant test, which are feasible in terms of the provisions of our study, and which are of some possible interest to those who are interested in the development of children.

In general, our criteria of development have been built to parallel the component parts of the infant examination procedure. It is suggested that the *adaptive* portion of the infant examination procedure might be most relevant to the measures of general intelligence provided by the Binet and the Arthur. The *fine motor* component of the infant-examination procedure might be relevant to the child's score on the Arthur, which is manual in implication, and with the psychologist's rating of the child's performance in manual respects. The *gross motor* portion of the infant examination procedure may seem to have a relevance for the gross motor ratings made by the pediatrician and the strength tests we have incorporated in the pediatric examination; in addition, this portion of the infant-examination procedure would conceivably be relevant to the structural development of the child as observed by the examining pediatrician. The *linguistic* aspects of the infant examination procedure would be relevant to the child's standing on the Stanford-Binet, which is a verbal measure of intelligence, the ratings of speech made during the examination, and by the (older) children's standings on reading tests and spelling tests. The *personal-social* aspect of the infant-examination procedure presumably would be relevant to a child's responses to the *Social Reaction Interview,* the manner in which he is described by his mother in the *Child-Rearing Practices Interview,* and the ratings of the child's responses in the testing situations made by the pediatrician and by the psychologists.

Perhaps these criteria would also be relevant to the informal characterizations of the child's personality which were made by the examiners and included in the informal description of the child.

Special neural or sensory evaluations based on the infant examination could be relevant to a child's over-all sensory competence as roughly estimated by the examining pediatrician, his neurological status as determined by the standard neurological examination, and by other details of the pediatrician's examination of a child's general development.

Thus it may be seen that we have sought in our small fraction of all conceivable criteria to sample kinds of behavior which would be relevant to the most conspicuous parts of the infant-examination procedure. In addition, the recommendations for the child with respect to adoptability are considered a part of the examination procedure. This, of course, can be validated in a sense by correlating it with the adoptive parent's reaction to the child.

It should be clear that a useful infant examination need not predict all these possible criteria. Any relationships found between the infant examination and the criteria could be an important reassurance to those who use an infant examination procedure. Nevertheless, if none of the aspects of the infant-examination procedure is found to have a useful degree of relationship with any of the criteria we have adopted, it would be very difficult to justify claims for the validity of the infant examination until criteria were devised which could be reliably predicted by some aspects of the infant examination. Specifically, if the infant examination does not predict any of the criteria we have adopted, then what of value does it predict? The cri-

teria that we have sought to employ, though partly of our own devising, are not esoteric criteria; they include the kinds of behavior which both child psychologists and parents understand and consider to be important.

Table 2 provides a summary of the most promising of the various criteria which we employed in our exploration of the predictive validity of the various facets of the infant examination. Parts III, IV, and Section C of Part V of our outline of criteria will be seen to have a specific representation in Table 2. This representation requires comment. The physical examination in particular, and, to a lesser degree, the procedure for the psychological examination, provide opportunities for numerous ratings and evaluations in addition to specific measurements employed. Some of the informal features of these procedures are doubtless inconsistently defined and possibly unreliably observed. In order to

TABLE 2

CRITERIA OF DEVELOPMENT SELECTED AS MOST PROMISING ON THE BASIS OF PRELIMINARY EXPLORATION

I. Intellectual Development (younger and older samples both)
  A. The Stanford-Binet
  B. The Arthur Point Performance Scale

II. Scholastic (older sample only)
  Reading
  Spelling
  Arithmetic

III. Motor Development (younger sample only)
  Physical Examination Items
    Dynamometer, L.H.
    Dynamometer, R.H.

IV. General Physical Development (younger sample only)
  Physical Examination Items
    Height quotient
    Weight quotient

V. Personal-Social Development (younger and older samples both)
  A. Child's Reaction to Himself (Social Reaction Interview)

Younger Sample:	Older Sample:
Dependence on Adults	"Goody-Goody"
Aggression	Responsible Attitude
Socialized Compliance	Cooperation with Authority
Takes Adult Role	
Weakness-Avoidance	
Constructive Approach	

  B. Mother's Reaction to the Child (Child-Rearing Practices Interview)

Younger Sample:	Older Sample:
Compulsive	Controlled
Aggressive	Spoiled child
Phobic	Dependent
Dependent	Anxious-aggressive

  C. Ratings Made by Examiners
    Test Reaction Scale Items (younger and older samples both)
      Sensitivity to failure
      Reaction to failure
      Belligerence
    Physical Examination Items (younger sample only)
      Approach to examiner
      Reaction to undressing
      Activity
      Composite pediatrician's rating

eliminate the less meaningful data which these procedures provide and in order to select out the more promising data, all the various items, ratings, and bits of information were correlated with all of the facets of the infant examination and were correlated with our various measures of the environment. The elements of our data which produced no relationship of important degree or fewer evidences of relationship than could be expected on the basis of chance alone were discarded, and the remaining more promising elements are included in Table 2.

As we have noted in other paragraphs, our criteria for development vary greatly not only in content and source but in certain important technical respects. The criteria for mental ability, the Binet and the Arthur, are well known, well standardized, reliable, and valid in the sense that their relationships with other variables are known. Similar advantages may be claimed for the scholastic achievement tests. Some of the aspects of the physical examination,[1] particularly the measures of size and strength, are well known and require little description or justification. Our measures for the social and personal aspects of children's development do not have all these advantages. Most of the cluster scores and the various ratings for personal and social development are novel and untried. Although we believe that little defense or justification for their purpose and content is required, it must be explicitly acknowledged that as measuring devices they are untried. We may suspect that as a group they are less reliable than our measures for mental ability. It is probable that some of these criteria are inappropriate for a study of the *degree* of relationship. Accordingly, we must not expect evidence of exact degree of relationship from these criteria. Much may be said in defense of these criteria, however. They meet certain practical requirements. As will be noted in our analyses, they are capable of yielding interesting and reliable correlations and accordingly, they obviously cannot be devoid of reliability. (In addition, we have shown promise of reliability for the cluster scores by the the use of the split-half method.) We believe, therefore, that if the infant examination has useful predictive validity for the behavioral content represented by these cluster scores, the cluster scores will be sufficient to show evidence of such validity. We find no reason for fearing that intrinsic validity for the content of these clusters will be obscured because of the imperfect nature of our measures.

Relationships Between Aspects of the Infant Examination and Criteria of Development

We sought evidence of relationship between all the various aspects of the infant examination and all of the criteria which were described in the preceding section. Our analyses involved certain intricacies and observed certain precautions that can be better illustrated than explained abstractly. Accordingly, we shall begin by a description of our study of the relationships between the General Maturity Quotient of the infant examination and Binet IQ at the time of our follow-up study. We begin with a discussion involving the General Maturity Quotient because this was the most promising aspect of the infant examination; almost all of the others showed no hint of validity.

It will be remembered that we expected to base our main analyses on the data for the younger group and to use the data for our older group to confirm or qualify the results of analyses for the

younger group. It should be remembered also from our discussion of the infant-examination procedure and from our discussion of our samples that all of our children do not lend themselves equally well to a validity study such as ours. Specifically, some of the children were placed adoptively *before* an infant examination. It seems highly improbable for those children that any relationship between the infant examination and subsequent criteria for development could be confounded with environmental differences resulting from selective placement based on the infant examination. We know that there were other children, however, who were placed adoptively *after* the infant examination, and that for these children there was some chance that the infant examination could have been a factor in the selective adoptive placement of the children. As a matter of fact, during the period when our children were placed adoptively, it was common practice in many southern New England agencies to depend upon the infant examination as an important aid in making decisions concerning the selective placement of adoptive children. Accordingly, we could suspect that the relationships between the infant examination and the criteria could be confounded by differences in environment emerging or organized on the basis of differences in infant test scores. More may be said in this connection.

It will be remembered that the examiners were not equally confident of the examinations for all of the children, that the examiners refused to classify some children on the basis of the first infant examination and suggested that the child be returned at some later time for another examination which might permit a more confident classification of the infant. It would seem possible, therefore, that the infant examinations in which the examiner had the greatest confidence and which led to the prompt classification of the infants would be more likely to influence the selective placement of the infants than examinations which in some respects were qualified and which the examiner did not wholly trust. In this connection it should also be recalled that those who advised us indicated that, in general, the examiners had more confidence in this examination when given between the twelfth and fortieth weeks than when given at other ages. It is possible, therefore, that examinations given to infants during this optimal age period may have been more important in influencing the selective placement of infants than examinations given at other ages. Accordingly, our analyses will distinguish between those infants who received their first examination at an optimal age and were confidently classified and those infants who either did not receive their first infant examination at an optimal age or who were not confidently classified as a result of the first infant examination.

Table 3 illustrates the distinctions that we have made among our children. This table summarizes the results of our validity analysis where the predictor variable is the General Maturity Quotient on the infant examination and the criterion is Binet IQ at the time of follow-up. (Incidentally, Table 3 summarizes our analysis between the General Maturity Quotient and the IQ on the Arthur Point Performance Scale at the time of our follow-up, too.) It may be noted that Table 3 describes our results for both the younger and the older groups.

Let us first consider the results for the younger group. Here the younger group is divided into an "after" group,

TABLE 3

RELATIONSHIPS BETWEEN THE GENERAL MATURITY QUOTIENT ON THE INFANT EXAMINATION AND MEASURES OF MENTAL ABILITY AT THE TIME OF THE FOLLOW-UP STUDY

Sample	With Binet IQ		With Arthur IQ	
	r	*N*	*r*	*N*
I. THE YOUNGER ADOPTIVE SAMPLE:				
A. Infants placed in adoptive homes *before* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	−.14	32	−.16	32
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks*	.15	21**	.13	21**
B. Infants placed in adoptive homes *after* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	.55	30	.19	30
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks	.16	30	.03	30
II. THE OLDER ADOPTIVE SAMPLE:				
A. Infants placed in adoptive homes *before* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	.20	16	−.06	16
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks	.42	7**	.62	7**
B. Infants placed in adoptive homes *after* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	.15	37	.28	35**
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks	−.02	18	−.22	18
III. THE NURSERY SCHOOL SAMPLE	−.14	35		

Note.—The average correlation between the General Maturity Quotient and the Binet IQ for all the various adoptive subsamples except the one which provides evidence for selective placement (the 30 younger adopted children in the selected after group) is .09. If this average includes the nursery school group also, it becomes .05. If an average correlation is computed which includes all the various subsamples except the younger one where selective placement was indicated, the nursery school group, and the selected younger sample when environmental differences have been partialed out of the correlation one finds the average correlation between the General Maturity Quotient and the Binet IQ to be .08.

* Two of these were examined before 12 weeks; one at 9 weeks, the other at 11 weeks of age.

** We did not have complete data for all cases; therefore, some of these correlations are based on *N*'s slightly smaller than the total *N*'s in the subsamples (see Table 1).

the group where the infants were placed adoptively after their infant examinations. A "before" group was also indicated, this was a group where the infant was placed adoptively before the infant examination. In the "after" group one could suspect selective adoptive placement occurring in part as a result of the infant's standing on the infant examination. In the "before" group one could have no basis for suspecting that the infant examination could have been a fac-

tor in selective adoptive placement. It will be noted that both the "after" and the "before" groups are divided into two parts. One part which we have selected for special emphasis comprises those whose first infant examination was administered at an optimal age and led to a confident classification of the infant. The other part comprises those whose first infant examination either was not given at an optimal age or was in some manner qualified and did not lead to a confident classification.

We would anticipate that *selective* adoptive placement based on the infant examination would be particularly probable for the selected "after" group of children who received their infant examination at an optimal age and whose infant examination led to a confident classification. It is obvious that we must be careful in our use of these "after" groups in our validity analyses. It is possible that such groups could yield seriously misleading evidence incorrectly suggesting a relationship between the infant examination and some criterion, e.g., the Binet. Such a confounding of the relationship could emerge because of selective arrangement of the relevant environment occurring as a result of selective adoptive placement based on the infant examination.

In view of our concern about the possible effects of selective placement, the data provided by Table 3 are most interesting. The only reliable relationship between the General Maturity Quotient and the Binet IQ occurred for the selected "after" group where selective placement based on the infant examination was considered the most probable. Specifically, the only correlation for the younger children which might possibly be used to refute the hypothesis of no relationship between the General Maturity Quotient and Binet IQ is for the very subgroup where selective adoptive placement (and the confounding with the environment) seems most probable. It is interesting to look at the data provided by the older sample. None of its subgroups provides any evidence for the refutation of the hypothesis of no relationship. To be sure, one of the very small older subgroups produces a high correlation. This correlation is not reliably greater than zero, however, and scrutiny of our data shows that it is almost wholly the result of two cases.

If for the younger sample one were to exclude from consideration the subgroup where selective placement seemed most probable, the selected "after" group, the average correlation for the three remaining subgroups would be only .05. If one were to take the average correlation among the four subgroups comprising the older sample, a value of .14 would emerge. Thus it would appear that if one were to discard the subsample where selective placement would seem most probable, we would have no basis for refuting the null hypothesis.

Since the probability of selective placement is so great for the selected "after" group, we do not consider it to be appropriate for our analyses and it should be excluded from consideration. Because we suggest that this subsample be excluded from consideration and because of our interest in adoptive placement practices as they may have involved the infant examination, it is of interest to examine the plausibility of these selective placement hypotheses.

In order to show that selective placement could have occurred in the particular subgroup where we had suspected it, we conducted two parallel studies of

possible relations between aspects of the infant examination and aspects of the child's environment. These parallel studies required two groups. One group comprised the 30 children who were tested at an optimal age, confidently classified on the basis of the first examination, and placed in adoptive homes *after* the first examination. The other group was also tested at an optimal age and confidently classified; they, however, (32 in number) were placed in adoptive homes *before* the first infant examination. Since the selected "after" group was for the most part agency placements and the selected "before" group was mostly independent placements, these two groups could have been somewhat different from each other with respect to infant-examination scores and measures for the environment. With respect to the infant examination, however, the differences between our selected "before" and selected "after" groups were not great. For example, in the selected "before" group (comprising 32 children) the mean and the standard deviation of the General Maturity Quotient on the first infant examination were 102.3 and 12.3, respectively. For the selected "after" group (comprising 30 children) the mean and the standard deviation on the first infant examination were 97.0 and 11.7, respectively. The difference between the means appears to be due to one extreme case. The difference between the standard deviations is not significant and accordingly we need not be concerned about differences in range of infant examination scores when we conduct parallel analyses for the selected "before" and selected "after" groups.

It is our particular interest to see whether or not there are some environmental differences which are predictable from the infant examination for the selected "after" group but not predictable from the infant examination for the selected "before" group. If the infant examination is used as a basis for selectively placing adoptive children in adoptive homes, we would expect correlations between some aspects of the infant examination and some aspects of the adoptive homes for groups of adoptive children who were placed by agencies *after* the infant examination. We would not expect such correlations to appear for groups of children who were placed adoptively *before* they received an infant examination.

Let us review some of the more interesting indications from this parallel analysis. Beginning with the *Ambition* cluster score for the home (see our report *Interviews as a Source of Scores for Children and Their Homes* for a discussion of environment scores), we find no evidence of important relationships for the "before" group. (The correlation between the General Maturity Quotient and the *Ambition* cluster score is -.17.) For the "after" group, however, we find indications of relationship between the General Maturity Quotient and the *Ambition* cluster score. (The respective correlation is .38.) Thus the "after" group not only provides a positive correlation but the correlation coefficients for the two groups are significantly different from each other.

There is no hint of important relationship between the *Harmonious* home cluster and the infant examination.

For the *No Ambition* cluster, however, there are indications of relationships. It is found that for the "after" group there is an inverse relationship between aspects of the infant examination and this cluster score; this suggests selective place-

ment. Curiously enough, there is a positive relationship between the *No Ambition* cluster score and the respective parts of the infant examination for the "before" group.

Continuing with our discussion of the cluster scores, we find no indication of relationship between status on the infant examination and the *Rejection* cluster score for the home. Similar apparent lack of relationship characterizes our survey for the *Parental Incompatibility* cluster score and the *Overprotective* cluster score. The balance of our analyses for the various cluster scores suggests that selective placement did occur for those children who were placed after their infant examination. One interesting instance is the postitive relationship between good standing on the infant examination and a tendency for adoptive parents to punish aggressive expression in the children ("after" group).

For these two groups of children we find certain additional evidence for selective placement. For the selected "after" group there is indication that status on the infant examination is related with the education-occupation status[7] of the adoptive father ($r = .47$). There is no such indication for the selected "before" group ($r = -.12$). These correlations are reliably different.

With respect to the difference between the selected "before" group and the selected "after" group in apparent degree of correlation between the General Maturity Quotient on the infant examination and the Binet IQ at time of follow-up, certain additional comments may be offered. One might expect that the difference between these two correlations is somehow or other an artifact of some insignificant differences between the two groups with respect to age at infant examination. In order to examine this possibility minutely, it was possible to arrange a selected "before" group and a selected "after" group so that they were exactly matched with respect to age at time of infant examination. For these respective matched groups of 22 cases, the correlation between General Maturity Quotient and Binet IQ was again determined and the two correlations were found to be reliably different.

In order to pursue the hypothesis that the correlation of .55 for the *selected "after" group* was an artifact of selective placement based on the infant examination, it was decided that we should attempt to consider the effect of probable aspects of the agencies' selective placement practices on this correlation. In the preceding paragraphs we indicated that we have provided evidence of selective placement with respect to the Ambition–No Ambition[8] aspects of the adoptive home, and with respect to the education-occupation status of the home. Our

[7] The composite education-occupation scale forms a continuum representing a combination of educational and occupational achievement of the home, and was based on the following operation: If the adoptive father's occupation was business or professional or in the white-collar group, a score of plus one was arbitrarily assigned. If, however, the adoptive father's occupation was unskilled, semiskilled, or skilled, the weighting was minus one. Weighting of zero was given in all situations where we did not know the occupation of the adoptive father. The father's education was combined by allowing plus one if he had attended college, minus one if he had not attended college, and zero when we did not have this information. The mother's education was scored in a similar manner. These three sets of weights were combined algebraically to form a general composite score representing the education-occupation status of the home.

[8] This Ambition-No Ambition composite cluster score was formed by giving the score on the *Ambition* cluster a plus sign and the score on the *No Ambition* cluster a minus sign, and then taking the algebraic sum of the two scores.

conversations with the various agency representatives as well as the published reports of other investigators (4) indicate that the age of the adoptive parents and the duration of their marriage was also taken into consideration. Accordingly, we formed an Age–Duration of Marriage composite[9] score.

If our selected "after" group had been very large indeed, it might have been possible for us to find subgroups which differed from each other with respect to the environmental factors resulting from selective placement, but which were homogeneous within themselves with respect to these environmental factors. An average correlation between the General Maturity Quotient and the Binet IQ for such subgroups of homogeneous environments would be a correlation free from the influence of differences in the environment resulting from selective placement. Our group is obviously too small to make possible such an analysis, but from the statistical standpoint the partial correlation coefficient is considered to be the equivalent of such an hypothetical average correlation.

Accordingly, a partial correlation between the General Maturity Quotient and the Binet IQ was computed for the selected "after" group. The partial correlation represents the correlation between the General Maturity Quotient and the Binet IQ after the influence of differences in the Ambition–No Ambition cluster score, the Age–Duration of Marriage score, and the Education-Occupation score have been eliminated from the original correlation. Under these circumstances the original correlation of .55 is reduced to .28. We consider this reduced correlation to be the best estimate of an average of correlations which might be found for hypothetical subgroups which differed from each other with respect to these environmental factors (seemingly organized on the basis of selective placement) but which within themselves were homogeneous in these environmental respects; we specifically regard this correlation as an estimate of an hypothetical correlation which is free from the aspects of confounding previously discussed.

Readers who prefer an hereditary explanation for individual differences may object to this partial correlation on the grounds that they consider both the differences in the infant examination and the differences in Binet IQ to be organized at least partly on the basis of common heredity and that any organization of the environment which occurred as a result of differences in the infant examination (or as a result of hereditary differences in the true parents) was in part at least an expression of the kind of hereditary factors which were responsible for the relationship between the General Maturity Quotient and the Binet IQ. Such readers could claim on the basis of their beliefs that our partial correlation analysis partialed out too much, i.e., it partialed out differences which were an expression of the relevant underlying heredity. In order to answer this question, we can offer two consider-

[9] The age-duration composite was prepared in the following manner: If the adoptive father were in his thirties, a score of plus one was arbitrarily given; if he were under thirty or forty or over, a score of minus one was given; if his age were unknown, a score of zero was given. The same scoring system was used for the age of the adoptive mother. In addition, if the couple had been married for seven or more years, a score of plus one was given; if they had been married for less than seven years, a score of minus one was given; and if the duration of their marriage was unknown to us, a score of zero was given. These three weights were added algebraically to give us a composite measure of the age of the adoptive parents and the duration of their marriage.

ations. First, the correlation between the infant examination and the education-occupation status of the true parents is only .16. This provides no basis for believing that the infant examination is an expression of any prepotent hereditary factors of the true parents which are sufficient to result in differences in their educational-occupational attainments.

Our more effective answer to this question involves the use of a control group. We were able to include in our follow-up study 35 children living with their true parents. These children had been examined in infancy as a part of preparations for their enrollment in the Clinic nursery school. Unfortunately, the correlation for this group was −.14. If an arrangement where a congruence between characteristics of the true parents and characteristics of the child's home environment could produce a correlation between the infant examination and criteria of development, we should most expect such a correlation to emerge for this group. This discouraging correlation of −.14 obviously fails to provide any indication that our partial correlation procedure disregards an important hereditary component in the correlation of .55 for our selected "after" group (where selective placement based on knowledge of either the true parents or the infant examination could have occurred). Incidentally, this sample of 35 children provides an additional instance where an hypothesis of no predictive value could have been refuted (if refutable by the use of our criteria). Unfortunately, this sample provides no challenge to the null hypothesis (the hypothesis of no relationship).

It should be remembered that we sought evidence of relationship between each of the various facets, scores, and indices of the infant examination and each of our various criteria for development. The Classification on the infant examination, which includes both the quantitative aspects of the infant examination and certain qualitative aspects, was found to reveal a pattern of relationships with the Binet IQ which was very similar to the pattern revealed by the General Maturity Quotient. The selected "after" group showed a correlation between Classification and the Binet IQ of .60. The emergence of this particular correlation is quite suggestive of selective placement based on the infant examination. Incidentally, the correlation for the selected "before" group was −.06. Certain preliminary analyses showed that the examiner's descriptions at the time of the infant examination ("drive and interest") was reliably related with Classification but not with General Maturity Quotient.[10] We had found that the children who were characterized by "sustained drive and interest" tended to be placed in those homes which eventually showed relatively high scores on punishing for aggression. When differences in the Punishment for Aggression cluster score, as well as differences in Ambition–No Ambition, Age–Duration of Marriage, and Education-Occupation were partialed out of the correlation between Classification on the first infant examination and Binet IQ at time of follow-up, the correlation was reduced to .30. In this connection the reader may wonder whether infants who were characterized by sustained drive and interest developed into children who required punishment for aggression. In order to answer this question, we examined the relationship between "sus-

[10] This is consistent with the composite basis for the Classification as contrasted with the psychometric basis for the General Maturity Quotient.

tained drive and interest" and "punishment for aggression" for our selected "after" group and our selected "before" group. For our selected "after" group the relationship was significant, whereas for the selected "before" group the relationship was not significant. This is taken as a further indication that drive and interest were factors in the selective placement of the infants and not necessarily a factor which predetermined the kind of child-rearing practices that would emerge as a consequence of a child's "intrinsic" personality.

The reader may be dissatisfied with our partial correlation analysis and we do not regard the results for the selected "after" group as first-rate evidence. This most obviously is a group where selective placement based on the infant examination is most probable and perhaps the evidence provided by this group should be excluded from consideration. It is of interest, however, to show that the indication of relationship between the infant examination and the Binet criterion of mental ability occurs for this group only and may be plausibly explained in terms of a selective placement hypothesis. Stated differently, by a consideration of the selective placement hypothesis, the indication for this most dubious subsample may be brought in line with the indications for all of the other subsamples.

As we terminate our discussion of the criteria for mental development, one may wonder why the selected "after" group with the strong evidence of selective placement based on the infant examination did not provide evidence of a correlation between the General Maturity Quotient and the Arthur IQ. The explanation for this may lie in the fact that the various environmental factors which the agencies seemingly considered in their selective placement practices are not shown by our analyses to be as consistently or as reliably related with the Arthur IQ as they are with the Binet IQ. From this one would infer that the differences in ability which the Binet represents and the differences in ability which the Arthur represents do not emerge in the same manner as a consequence of environmental differences. Our numerous analyses involving all facets of the infant examination and the criteria for mental development provide no basis for refuting the hypothesis that there is no relationship between the infant examination and the criteria for mental development. In view of all the relationships examined, fewer promising trends emerged than could be accounted for on the basis of chance alone, and these promising trends were found only for a subsample where selective placement based on the infant examination was not only anticipated on a priori grounds but was indicated on the basis of analyses. In addition, partial correlation analyses sustained and illuminated the selective placement hypothesis for the emergence of these few correlations. Despite the fact that we used numerous subgroups, two different age samples, and a sample of children living in their own homes, we could find no challenge to the hypothesis of no relationship, and the implications of our study are congruent with the implications of others who have examined this question (1, 3, 5).

Our data for the older sample included scholastic achievement criteria. There was no indication that any of these criteria was reliably related with any aspect of the infant examination.

When we consider certain "nonpsychological" aspects of the infant's development, we find some indication that the

infant examination may have some predictive value. Specifically, it is found that the adaptive developmental quotient on the infant examination has a correlation of .45 with strength of grip in the left hand as measured by the dynamometer at the time of our follow-up examination. This correlation of .45 is found for the selected "before" group, a group in which we have no basis for expecting selective placement or confounding from any other source. Although this correlation is plausible in the sense that the adaptive developmental quotient involved the use of the hands, the correlation was not reliable for the other subgroups and could be an accident of sampling. Since the physical examination was not conducted with the older sample, we could not explore this relationship further.

The weight quotient (a weight/age ratio) showed an interesting correlation with the fine motor developmental quotient of the infant examination. The correlation for the selected "before" group was .44 and for the selected "after" group it was .20. The other correlations were less promising.

It is interesting to note that the most definite suggestions of predictive validity of the infant examination are the physical aspects of development. It may be remembered from our discussion of the infant examination that this was an aspect of development which received relatively little emphasis in the literature concerning the infant examination (an examination for infant behavior). We should not overgeneralize, however, and conclude that our data show definite evidence that the infant examination has predictive value for these and other physical aspects of development. Our data are suggestive and should be followed by further inquiry where the physical aspects of development are the focus of the investigation.

The various possible relationships between aspects of the infant examination and our various measures for the personal and social development of the children were examined for our different subgroups. The many correlations were small and insignificant. From the numerous possibilities only two correlations of any interest emerged and both of these were for the selected "after" group which will be remembered as being characterized by selective adoptive placement based on the infant examination. For the selected "after" group the psychologist's rating of Sensitivity to Failure (made at the time of our follow-up study) was correlated to the order of .37 with the Personal-Social developmental quotient of the infant examination. The corresponding correlation for the selected "before" group was .19. In view of our a priori considerations and the results of our analyses with the Binet IQ, we are not surprised to find that this correlation of .37 for the selected "after" group was reduced to the order of .24 when differences in Ambition–No Ambition and Age–Duration of Marriage were eliminated from this correlation by the partial correlation device.

A similar pattern of relationships was observed for a special composite score which we prepared, a score comprising the child's approach to the pediatric examiner, his reaction to undressing, interest, affect, and activity at the time of the physical examination during the follow-up study. This composite score was correlated with the Personal-Social developmental quotient on the infant examination to the order of .31 for the selected "after" group. The correlation for the selected "before" group was −.10. When differences in the Ambition–No

Ambition composite score were eliminated from the correlation of .31 by the partial correlation device, the correlation was reduced to .27. It is probable that this correlation could have been further reduced by the use of other variables.

It should be noted that no promising correlations emerged in the selected "after" subsample for our *older* children. We examined the data for the older children and learned that the correlations between aspects of the infant examination and aspects of the environment were not as high as the respective correlations of the younger sample. Although the difference between the correlations for the two age groups would not have reached the criterion for statistical significance, the difference in magnitude may be sufficient to account for the fact that the correlations between the infant examination and the criteria did not emerge. Some of our subsequent analyses suggest an explanation. In general, the standards for selecting the adoptive homes were higher prior to 1945 when the older sample was being placed than subsequent to 1945 when the younger sample was being placed. The difference, particularly with respect to such variables as ambition (which, incidentally, was not a factor in the selective placement of the older children) were not differences in range but differences in level. It seems possible, therefore, that some of the conspicuously less favored homes which might have had a detracting effect on the child's development were not represented in our older sample to an appreciable degree. This is not a necessary interpretation, however. It is possible that the *school child* is not dependent in his development on the home in the same manner as the *preschool child*. In this connection, however, it may be remarked that both the younger and the older sample provide evidence of relationship between aspects of the environment and aspects of development. More will be said of this in our report on the environmental relationships.

Conclusion

In conclusion, our exploration of the relationships between the numerous facets of the infant examination and the numerous criteria for development which we have employed has yielded a lengthy and monotonous series of correlations which were not only unreliable but which were too small to be of any practical interest had they been based on huge samples capable of making practically insignificant correlation coefficients statistically significant. The correlations between aspects of the infant examination and our criteria which were statistically significant and large enough to suggest some practical value were so infrequent as possibly to have occurred as a result of chance alone. Moreover, these more interesting correlations occurred in a subsample where selective placement on the basis of the infant examination was not only anticipated but was consistent with correlations between the infant examination and certain conspicuous characteristics of the adoptive parents which could have been known to the placement agencies. Partial correlation analyses sustained the selective placement hypothesis for this questionable group.

Aside from the few correlations produced by the subsample where selective placement based on the infant examination was conspicuous, the other interesting correlations were for strength of grip and weight quotient. These correlations were not found for all subsamples, and since the physical examination was not conducted for our older group, they could not be confirmed. They are far

from conclusive and only suggestive, but their implications should be explored by analyses which emphasize physical aspects of development.

Our investigations have sought evidence of predictive validity both with respect to facets of the infant examination which other investigators have not examined and with respect to criteria which other investigators have not employed. (Some of our criteria, such as ratings and cluster scores, are not as well tried or as reliable as our mental ability, scholastic achievement, or physical development criteria. Nevertheless, our cluster scores, and by inferences from numerous significant correlations, our ratings, too, offer a modicum of reliability and should be sufficient to indicate the presence of predictive validity despite the possibility that they would not yield suitably exact indications of *degree* of validity.) Despite the broad basis of our study, we, like the others, have no basis for challenging the hypothesis that the infant examination is without predictive validity. As we have frequently emphasized, however, such a result does not *prove* that the infant examination has no predictive validity, and one must remember that it is never possible to prove a hypothesis of no relationship.

Appendix A

TEST REACTION SCALE

Mother-Child Relationship as Observed by Examiner

1. Child was not at all hesitant about leaving mother, scarcely said goodbye and went with examiner readily. He showed no concern for mother during examination and was extremely casual about seeing his mother again at close of examination.
2. Child was somewhat hesitant at leaving mother, said goodbye and looked to see where she would wait, and showed pleasure at seeing her again at close of examination.
3. Child was extremely reluctant to leave mother, held on to her hand, may have cried, wanted her to come with him, needed to be reassured that she would wait for him. He asked about her several times during the examination. He seemed very much relieved to see her again.

Attention

1. Completely absorbed by task with no interference from distracting stimuli.
2. Almost always absorbed by task but a little interference from distracting stimuli.
3. Pays attention to task. Efficiency is not impaired by attention to outside stimuli.
4. Child is easily distracted by extraneous stimuli or by his own ideas, but returns readily to the task.
5. It is difficult to get and maintain the attention of the child.

Activity

1. Child can't sit still for any length of time. Often stands up, explores the room, shifts position, climbs around the chair.
2. Child generally sits still for the interview but may occasionally become restless, shift position, slide down in the chair.
3. Child sits still during most of the interview unless told to stretch or stand.

Sociability

1. Child adjusted to the examination with immediate friendliness. He smiles, laughs, shares experiences with the examiner shortly after meeting.
2. It requires a while for him "to warm up" to the examiner.
3. The child maintains an attitude of reserve for a considerable length of time. It is only toward the close of the examination that any friendliness is manifested.
4. The child at no time becomes friendly, laughs, smiles or talks spontaneously or freely with the examiner.

Belligerence

1. Child is markedly belligerent. He attempts to strike or insult the examiner or to mar equipment.
2. Child is surly and may deliberately refuse to cooperate.
3. Child displays no evidence of hostile reactions during examination.

Maintenance

1. After the child feels at ease with the examiner, favorable cooperation and motivation are sustained for the duration of the contact.
2. After the child feels at ease with the examiner favorable cooperation and motivation are generally sustained but the child may have moments of apprehensions, distrust or belligerence.
3. After the child feels at ease with the examiner favorable cooperation and motivation deteriorate so that the child cries, refuses to continue, becomes excessively fatigued.

Reaction to Failure

1. Child reacts to a feeling of failure with marked increase in effort and persistence.
2. Child reacts to a feeling of failure with a slight increase in effort and persistence.
3. Child reacts to a feeling of failure with slight decrease in effort and persistence.
4. Child reacts to a feeling of failure with a marked decrease in effort and persistence.

Sensitivity to Failure

1. The child is oblivious to his failures and indifferent to any suggestion of error.
2. The child shows some recognition of his errors and slight concern about errors.
3. The child is markedly sensitive to failure and shows considerable concern about errors.

Male Versus Female Examiner

1. Child exhibits a greater degree of enthusiasm, friendliness and confidence with a male examiner than with a female examiner when other factors remain the same.
2. Child exhibits the same degree of enthusiasm, friendliness and confidence with a male examiner as with a female examiner when other factors remain the same.
3. Child exhibits a lesser degree of enthusiasm, friendliness and confidence with a male examiner than with a female examiner when other factors remain the same.

Modesty

1. The child removes his clothes readily.
2. The child removes his clothes only with encouragement and persuasion.
3. The child is extremely reluctant to remove his clothes.

Thumb Sucking

1. During the time the child is observed at the clinic he sucks his thumb whenever otherwise not engaged.
2. During the time the child is observed at the clinic he sucks his thumb intermittently.
3. During the time the child is at the clinic he is observed to suck his thumb.
4. During the time the child is observed at the clinic there is no thumb sucking.

Fine Motor Ability

1. The child exhibits continuous and precise controls so that fine motor activities are rapid and accurate.
2. The child exhibits some control over fine motor activities but movements are occasionally inaccurate or somewhat slow.
3. The child exhibits little control over fine motor activities so that movements are inaccurate or clumsy.

Speech Defect

1. The child has a sufficiently severe speech defect that interferes with communication.
2. The child has a mild speech defect which, however, does not interfere with communication.
3. There is no evidence of speech defect.

Appendix B

A.C.R.P. EXAMINATION FORM

Name _____ Sex _____ Birth Date _____ Date of Ex. _____

Height _____ Weight _____

Pertinent History:

Head: Circumference _____

Shape: Dolichocephalic _____
Brachycephalic _____
Other _____

Hair: Color _____ Consistency _____

Eyes: Color _____ E.O.M. _____ L.&A. _____ Fundus _____
Vision: R L
Other

Ears: External _____ T.M. _____
Hearing: R L
W.V. (15″)

Nose:

Mouth: Teeth: Number _____ Caries _____ Fillings _____
Orthodontic _____
Tongue:
Mucosa:

Throat: Tonsils: Present _____ Absent _____
Size _____

Neck: Adenopathy:
Thyroid:

Chest: Shape:
Lungs:
Heart & Circulation:
Rate _____
Rhythm _____
Murmurs _____
Size _____
Femoral Pulse _____

Abdomen: Shape:
Liver: Kidney: Spleen:
Tenderness:
Masses:

Genitalia: Circumcision: Other:
Testes Descended:

Skin:

Ext. Back Spine _____ Deformity _____

Joints _____ Limitation of Motion _____

Neurological:

Cranial Nerves _____
Motor Nerves _____
Sensory Nerves _____

Reflexes: 1. Hyper 2. Normal 3. Hypo 4. Absent

	R	L
T		
B		
BR		
Abdomen:		
Upper		
Lower		
K		
A		
P		
Crem.		

Gross Motor:

Agile
Normal
Clumsy

Handedness: R _____ L _____

Strength: Dyn. _____
Jump _____
Pull _____

Personality

Approach to Examiner

_____ 1. Spontaneous
_____ 2. With suggestion
_____ 3. With urging
_____ 4. Resistive

Speech

_____ Spontaneous
_____ Hesitant
_____ Lisping
_____ Stammer
_____ Infantile
_____ Other

Reaction to Undressing

_____ Unassisted
_____ Partly Assisted
_____ Assisted
_____ Prompt without show of concern

_____ Cooperative but not enthusiastic
_____ Great resistance

General Description:

Interest and Curiosity:
Affect:
Activity:

General Impression:

REFERENCES

1. ANDERSON, L. D. The predictive value of infancy tests in relation to intelligence at five years. *Child Develpm.*, 1939, **10**, 203-212.
2. ARTHUR, GRACE. *A point scale of the performance tests*, (2nd ed.) New York: Commonwealth Fund, 1943.
3. BAYLEY, NANCY. Mental growth during the first three years. *Genet. Psychol. Monogr.*, 1933, **14**, 1-92.
4. BRENNER, RUTH F., *et al. A follow-up study of adoptive families*. New York: Child Adoption Research Committee, Inc., 1951.
5. CATTELL, PSYCHE. *The measurement of intelligence of infants and young children*. Lancaster: Science Press, 1940.
6. GESELL, A. The stability of mental growth careers. *Yearb. Nat. Soc. Stud. Educ., Part II*, 1940, **23**, 149-160.
7. GESELL, A., & AMATRUDA, CATHERINE S. *Developmental diagnosis*. New York: Hoeber, 1941.
8. GESELL, A., THOMPSON, HELEN & AMATRUDA, CATHERINE S. *The psychology of early growth*. New York, Macmillan, 1938.
9. GRAY, W. S., *Gray's Oral Reading Test*. Bloomington, Ill.: Public School Publishing Company, 1915.
10. ILG, FRANCES L., & AMES, LOUISE B. Child behavior. *New Haven Register*, Feb. 3-5, 1953.
11. KELLEY, T. L., RUCH, G. M., & TERMAN, L. M. Stanford Achievement Test—Form D. Chicago: World Book Company, 1940.
12. TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.

(Accepted for early publication October 18, 1955)